

5/petB

10/534538

JC20 Rec'd PCT/PTO 12 MAY 2005

The polynucleotide sequences encoding full length of chicken type II collagen and use thereof

FIELD OF THE INVENTION

The present invention relates to a nucleic acid molecule as set forth in SEQ ID NO:1 comprising a polynucleotide sequence encoding full length chicken type II collagen (CCII), or a fragment thereof in possession of the same biological functions as well as CCII encoded thereof. It also relates to a method for producing CCII, and its use in the manufacture of a medicament for treating and/or preventing rheumatoid arthritis (RA). The invention specifically relates to a pharmaceutical composition for treating and/or preventing osteoarthritis and RA, to a food or beverage composition, and to a food additive composition, containing CCII prepared according to the method described in this invention, and the use of the nucleic acid molecules of the present application in gene therapy.

BACKGROUND OF THE INVENTION

RA, an autoimmune disease with a high incidence, can seriously threaten the health of the afflicted individual. Due to the limited knowledge concerning the etiology and pathogenesis of RA, controlling inflammation, alleviating the symptoms, and maintaining the functions of affected joints are the principal therapeutic aims when treating those diagnosed with RA. However, these treatments are far from the ultimate goal of preventing joint injury. With the development of novel molecular biology techniques in recent years, great progress has been made in our understanding of the pathogenesis of RA, providing inroads into new therapeutic strategies.

In 1993, American scientists reported in for the first time their ability to successfully treat RA patients by using CCII (Trentham DE, Dynesius-Trentham RA, Orav EJ, et al. Science, 1993,261: 1727-1730), a finding which immediately attracted world-wide attention. Phase III clinical trials with CCII are underway in some developed countries, including America, Britain, and France. From a comprehensive analysis of the published literature we

concluded that an oral tolerance approach, as a novel treatment, may increase our prospects for treating RA, by allowing investigation into both its root cause and symptoms (Trentham DE, Dynesius-Trentham RA, Orav EJ, et al., Science, 1993, 261: 1727-1730).

To date, these clinical trials are only half-way toward completion and the pharmaceutical companies in developed countries are beginning to make CCII into an effective food additive to circumvent the laborious medication approval process. Though, internationally, natural CCII has been administered in studies of oral immunological tolerance for RA, its use in this manner also has severe drawbacks: 1) The quality of the obtained CCII differs from one company to another, as well as between batches from the same company, which makes it difficult to guarantee a consistent and continuous curative effect; 2) The continued extraction and preparation of CCII is time-consuming and requires a good deal of resources.

The inventor proposes a new approach, specifically a genetic engineering method for producing CCII, to solve the above-mentioned problems. We developed a technique to clone the CCOL2A1 gene, which can then be expressed efficiently in order to produce the recombinant CCII. The proposed invention relies heavily on the successful novel cloning of the polynucleotide sequence -CCOL2A1 cDNA encoding the full length CCII.

SUMMARY OF THE INVENTION

In one aspect, the present invention relates to an isolated nucleic acid molecule as set forth in SEQ ID NO:1, comprising a polynucleotide sequence encoding the full length CCII, or a fragment thereof in possession of the same biological functions.

In another aspect, the present invention also relates to CCII or the fragment thereof in possession of the same biological activity, which is encoded by the isolated nucleic acid molecule of the above-mentioned present application.

Additionally, the present invention relates to a recombinant expression vector containing the nucleic acid molecules encoding CCII, or a fragment thereof in possession of the same biological functions.

The present invention further includes the host cells transformed with the above-mentioned recombinant expression vectors, which is able to express CCII, or a fragment thereof in possession of the same biological activity.

The proposed invention further includes a method for producing CCII, comprising the following steps:

- 1) Transforming a suitable host cell by the recombinant expression vector with the present invention;
- 2). Culturing the host cells obtained in step 1) in suitable culture medium and under appropriate culture conditions;
- 3). Separating and purifying the proteins of interest from the culture medium or cell.

The present invention further relates to the use of CCII prepared according to the proposed invention in manufacture of a medication for treating and/or preventing RA.

The present invention is particularly relates to a pharmaceutical composition for treating and/or preventing osteoarthritis and RA, which contains a therapeutically effective amount of CCII prepared according to the methods in the present invention, and optionally, a pharmaceutically-accepted vehicle.

The present invention also includes a food or beverage composition, which contains a necessary amount of CCII prepared according to the method described in the present invention.

The present invention also relates to a food additive composition containing a certain amount of CCII prepared according to the above-mentioned method of present invention.

The present invention further describes the use of the said nucleic acid molecules and fragments thereof in gene therapy.

DETAILED DESCRIPTION OF THE INVENTION

Recently, oral tolerance, a specific immunotherapy, has become the most widely studied treatment regimen in treating autoimmune diseases. In general, the process includes oral administration of a protein antigen, which is subsequently used for parenteral immunization and which then arouses a mild immune response. It is believed that the development of oral tolerance is mediated by several factors, including self-cytotoxicity, clonal anergy and deletion, and bystander suppression. Additional factors that determine the efficacy of the treatment include the dosage, origin, properties of the antigens, the present process of antigen, and the heredity and development of the individual receiving treatment.

We may therefore, conclude from a comprehensive analysis of the published literature that: 1) oral tolerance opens up a novel and promising way to treat RA by examining both its etiology and symptom profile; 2) oral tolerance of CCII produces a marked curative effect in many RA patients; 3) CCII is superior in its curative effects to other species-derived CII, because CCII is rich in chondroitin sulfate A and proteoglycans. The latter has the highest concentration of glucosamine sulfate, which possesses a powerful anti-inflammatory action and the ability to repair cartilage. In particular, 6- to 8-week-old chickens have the highest glucosamine sulfate; 4) CCII is also rich in proteoglycans with anti-oxidant activity (also known as cartilage matrix glycoprotein, CMGP), which reduces oxidation in chondrocytes; 5) CCII effectively prevents the proteasomal digestion of articular cartilage, and restores the damaged chondrocytes and cytokines, thus remarkably lessening the incidence of inflammation; 6) CCII promotes the synthesis of both chondrocytes and proteoglycans, and also increases secretion of synovial fluid and hyaluronic acid in joints; 7) CCII is a potent anti-inflammatory agent and pain reliever; 8) CCII is safe and free of any toxicity or side

effects, particularly when compared to any other RA drugs currently available.

At present, the immunological tolerance induced by oral administration of CCII has become an important novel treatment for RA. In order to provide sufficient amounts of high-quality CCII, we employed genetic engineering, instead of the traditional extraction and purification from the natural materials, to produce recombinant CCII.

The proposed invention involved the cloning of the CCOL2A1 gene, which was then expressed to produce recombinant CCII. In addition to the successful expression, this approach concurrently deepened our understanding of the CCOL2A1 gene. To date, human, canine and rat COL2A1 genes have been successfully cloned and localized on the chromosome, but studies that have cloned the full length CCOL2A1 cDNA and genome DNA, identified the location on the chromosome, described constitutive expression in various tissues, and induced expression of recombinant CCOL2A1 remain unavailable. These reports regarding the cloning of CCOL2A1 can provide great medicinal, therapeutic and economic value.

The cDNA sequence encoding CCII is relatively long, moreover, our understating about the functional region of the gene is greatly limited, thus it is very important to obtain the cDNA encoding the triple helical region of CCII. The application of PCR to isolate and amplify this gene is complicated in that CII has a complicated secondary structure, highly repetitive sequences, and a high GC content (mean GC content > 70%, some as high as 80%) (Nah DH, Upholt WB., J Biol Chem, 1991, 266 34:23446-23452).

Therefore, the restriction site of the cDNA sequence in the CCI3' domain was analyzed, in which the cDNA was incompletely digested by the restriction enzyme, followed by a PCR amplification of the C-terminal domain of the digested cDNA. Because of the long cDNA sequence encoding CCII, we divided the cDNA into five fragments, and each fragment underwent a separate PCR amplification with overlap of at least 50bp, to allow for a better end-to-end connection using SOE-PCR. Because of the high content of GC in genes

encoding CCII, Taq enzyme, which is particularly suitable for the amplification of GC-rich content, was used.

The obtained full length CCOL2A1 cDNA was 4837bp and consisted of an open reading frame (ORF) of 4260bp and a 3' nontranslated region of 520bp. The cDNA was inserted into a pGEM-T vector for sequencing, the results of which confirmed that the cDNA sequence was specific to CCOL2A1, which has thus far not been reported. The sequence has been admitted to GenBank database (AY046949).

The available literature indicates that the structure of the CCII genomic DNA is rather complicated, yet, to date no relevant studies on the CCOL2A1 genome DNA have explored this in detail. Therefore, following the successful cloning of the full length CCOL2A1 cDNA, we cloned and analyzed the CCOL2A1 genomic DNA. Because it contains a high content of GC, and frequently repeated sequences, as well as many nonspecific binding sites to the primers, we initially only obtained fragments (5494bp) of the 3' end from the amplification of the chicken peripheral blood corpuscles-derived DNA through PCR (anti-coagulating with EDTA), but failed to obtain the gene of interest, regardless of the template, primer, or PCR amplification conditions utilized. Furthermore, we also failed to sequence the positive clones screened from the chicken DNA library due to the high GC content and PolyT structure.

In view of this, we next designed several primers for sequencing based on the known exon sequence, and succeed in sequencing the DNA of the CCOL2A1 genome. The obtained DNA sequence was 12003bp long, contained 45 exons interrupted by 44 introns, which, as confirmed by a Genbank search, was unique to the CCOL2A1 genome, and had never been previously reported. The identified sequence has already been submitted to GenBank (Accession No: AF452711)

The sequence analysis of the CCOL2A1 genomic DNA clearly revealed that either the intron or the exon is remarkably smaller than in chicken type I and III collagen, consistent with the

reports by Upholt (Ausar SF, Beltramo DM, Castagna LF, et al., *Rheumatol Int.*, 2001, 20 :138-144); CII is highly conserved throughout evolution, its amino acid is 92% homologous in various species, and in some is 99% homologous. Based on this, we held that a short mean length of the introns, rather than a reduction in the overall number of introns, accounted for the compact CCOL21 gene.

To better understand the related genetic bioinformatics of CCOL2A1, and to gain insight into the evolutionary characteristics, homology was analyzed between the sequence of the full length cDNA of the cloned CCOL2A1 and the encoded proteins, and that of the triple helical region cDNA and proteins from dog (AF023169, AF242201), human (L10347), speckled dace (U23822), and rat (M65161), respectively, using MegAlign from Dnastar software. From this, the evolutionary trees were identified via Genedoc. The results revealed that CCII was most homologous to the canine CII (79.03 % and 94.77 % for the cDNA and amino acid, respectively), followed by the human CII (78.96% and 93.89%, respectively), and the rat CII (77.38% and 92.90%, respectively). However, among the five species that had been compared, human and canine CII share the greatest degree of homology (91.89% and 98.52% for the nucleotide and amino acid, respectively), with rat CII sharing the second greatest degree of homology (88.63% and 96.15%, respectively).

Following identification of the CCOL2A1 cDNA and genomic DNA, as well as the evolutionary characteristics of the gene, the next aim was an accurate localization of the CCOL2A1 gene on the chromosome. RH hybridization mapping, though recognized as the most precise method to localize a gene on its chromosome, was not possible with the CCOL2A1 gene as no chicken RH hybridization template is available from any suppliers. Thus chromosomes during the cell-division phase, derived from chicken peripheral blood corpuscles, were hybridized using our chromosome FISH technique, with digoxin-labeled gene fragments serving as the probes. As confirmed by the statistical analysis, genes encoding CCOL2A1 were located within the second region of the short arm of the chromosome 4 in chickens, and in chromosome 8 and 12 in rat and human, respectively. (Barnett ML,

Combitchi D, Trentham DE. *Arthritis & Rheumatism*, 1996, 39 4:623-628)

It is widely known that collagens, with the exception of Type XI and II collagens, are broadly distributed throughout various organs, and usually share common chains. Type XI collagen contains polypeptide subunit $\alpha 1$, $\alpha 2$ and $\alpha 3$, the last being the product of $\alpha 1(\text{II})$ gene expression; therefore $\alpha 3$ differs from the $\alpha 1$ of type II collagen only in the content of hydroxyl lysine (Rousseau JC, Farjanel J, Boutillon MM, et al. , *J Biol Chem*, 1996,271(39): 23743-8). Type XIII collagen is a transmembrane protein of C II, so they are constituted of similar polypeptide residues (Snellman A,Keranen MR, Hagg PO, et al. *J Biol Chem*, 2000, 275(12):8936-44).

To understand the important role of CCOL2A1 in the evolution and development of the chicken embryo, RT-PCR and ELISA were employed to analyze comprehensively and systematically the expression of CCOL2A1 in both the embryos and mature tissues. Our results suggest that in embryos, CCOL2A1 mRNA is expressed in the heart, liver, vitreous body, cornea, skin, pectoral muscle, sternum, small intestine, arthroidal cartilage, meniscus, and skull, but not expressed in the spleen, thymus gland nor testicle. Detection of CCII protein levels in adult chickens reveals that, in addition to expression in the the sternum and arthroidal cartilage, CCII proteins are also expressed in both the spleen and small intestine; the highest level of expression is found in the arthroidal cartilage, followed by the sternum. However, it is important to note that in the process of detecting CCII expression, when digesting with pepsin and elastase, CII can generate an α chain short peptide. Because of this, collagen containing $\alpha 1$ (II) chains may bind with the monoclonal antibody of CCII, generating false-positive labeling. As a result, the possibility that we simultaneously detected type XI and XII collagen was not excluded.

It should be emphasized that the splicing of the CCOL2A1 exons can be quite complicated, particularly the CII exon 2, as its splicing varies in different tissues. We therefore analyzed the splicing of exon 2 at the N-terminal domain from the cartilage tissues with high CII expression, including sternum and arthroidal cartilage, as well as non-cartilage tissues, such as heart, liver, pectoral muscle, small intestine, vitreous body, and cornea. In 17 day chickens,

exon 2 was absent from the cDNA encoding CCII in sternum, meniscus, and articular cartilage, but present in non-cartilage tissues, such as heart, liver, vitreous body, cornea, small intestine, muscle, skin and pectoral muscle. However, RT-PCR results indicated that exon 2 mRNA may be absent from the vitreous body and cornea. Furthermore, the vitreous body from the embryos <14d contain only the second exon, which was notably absent at day 17 of embryological development. This is the first report of differential splicing of the exon 2 in various tissues.

The above-mentioned phenomena may be explained in that the second exon of CCII is absent in the cartilage tissues, but present in the non-cartilage tissues, such as the heart, liver, cornea, small intestine, muscle, and skin. In addition, regulatory sequences exist in CII collagen that can influence expressions of CII in different tissues and in different embryological stages as well. CCII, initially distributed extensively throughout the internal organs, is ultimately restricted only to cartilaginous tissues during chicken growth and development. Finally, abundant CCII is found in the sternum from the developing chicken embryos, which is gradually reduced due to the calcification of the cartilage (Young MF, Vogeli G, Nunez AM, et al., *Nucleic Acids Res*, 1984,12 (10): 4207-4228; Marshall GE, Konstas AGP, Lee WR., *BrJ Ophthalmol*, 1993,77:515-524). Present views support the notion that CCII is expressed in both the vitreous body and in cartilage (Seery CM, Davison PF., *Invest Ophthalmol Vis Sci*,1991,32:1540-1550; Huerre-Jeanpierre C, Mattei MG, Weil D, et al., *Am J Hum Genet*, 1986, 38(1): 26-37), but this view conflicts with our findings that CCII is undetectable in the vitreous body.

To achieve the aim of the present invention, we succeeded in cloning the cDNA encoding CCII, and localized this gene to the chromosome. We also analyzed the gene expression profile, cloning of the genomic DNA, and the bioinformatics of CCOL2A1. In particular, the successful cloning of the cDNA encoding full-length CCII is of great importance as it provides a solid foundation for the proposed invention, allowing us to produce recombinant CCII and subsequent novel therapeutic agents to treat RA, using

genetic engineering approaches.

The present invention relates to the full length polynucleotide sequence comprising the nucleotides of SEQ ID NO: 1 encoding CCII, and/or the fragment in possession of the same biological functions.

The present invention also relates to CCII encoded by the isolated nucleic acid molecules, and/or the resulting fragment with the same biological functions, or a conserved variant thereof.

The "variant" of protein or polynucleotide refers to an amino acid sequence or polynucleotide sequence encoding the same having one or more amino acids or nucleotide modified. The modification here includes the deletion, insertion, or substitution of an amino acid or nucleotide. The variant may have conserved modifications in which the original sequence is replaced by another one of similar structure and chemical properties. Such an alteration may include, for example, substitution of leucine for isoleucine, or non-conserved changes, for example, a substitution of tryptophan for glycine.

"deletion" indicates one or more of the amino acids or nucleotides in the amino acid sequence or nucleotide sequence is deleted.

"Insertion" indicates the addition of one or more of the amino acids or nucleotides to the original amino acid or the nucleotide sequence.

"Substitution" indicates that one or more of the amino acids or nucleotides are replaced with various amino acid or nucleotide.

"Bioactivity" indicates that the proteins have the same structure, regulative or biochemical functions as the natural molecules;

"Similarity" indicates that the amino acid residue in the corresponding position should be the same, or conserved to a similar degree. The amino acid used in the conserved substitution includes: negatively-charged amino acids including aspartic acid and glutamic acid; positively-charged amino acids including lysine and arginine; amino acids with no electric charge and similar hydrophily to the head group, including leucine, isoleucine and valine; glycine and alanine; asparagine and glutamine; serine and threonine; phenylalanine and tyrosine;

"isolated" indicates the removal of the substance from its original environment, for example, a natural polynucleotide or polypeptide in the live animal is not separated. Similarly, a polynucleotide and polypeptide can be separated from some or all of its coexistent substances in the natural system. Such polynucleotide and polypeptide may be a part of the vector or the composition, which is also isolated because they are not a component of the natural environment.

The polynucleotide of the present invention may present as RNA or DNA, the latter includes cDNA, genomic DNA, and synthesized DNA. DNA can be double chains or the single chain (sense strand and/or antisense strand). The polynucleotide sequence encoding a polypeptide may be the same as in SEQ ID No.1, or a different sequence due to the redundant genetic code, which encodes the mature polypeptide in SEQ ID No.2.

The present invention further relates to a variant of the polynucleotide encoding the amino acid sequence, polypeptide fragment, and analogue of the SEQ ID No.2. These include polynucleotide variants, natural allele variants, or unnatural polynucleotide variants.

The polynucleotide encoding the invented polypeptide was fused with a specific flag sequence in the same reading frame, which can promote purification of the polypeptide. When using a bacteria host, the flag sequence may be a histidine marker of the pQE vector, so as to promote the purification of the polypeptide fused with the marker, the flag sequence

may also be a haemagglutinin (HA) marker from a mammalian host (such as COS-7 cell line of the monkey renal fibroblast). In addition, the polynucleotide sequences encoding the invented polypeptide may also consist of specific homologous or heterogenous sequences for the signal peptide, which can help with secretion of the proteins of interest out of the procaryotic cell or the eukaryotic cell membrane. The technicians involved in this project are aware that the above-mentioned flag sequence and the signal peptide sequence can be added to the vector expressing the polypeptide by recombinant methods or by a chemical process.

The present invention concerns the polypeptide encoding the amino acid sequence of the SEQ ID No.2, the activate fragment thereof, analogue, and variant.

The present invention relates to the polypeptide amino acid sequence of SEQ ID No.2 and the active fragment, analogue and deivant. The fragment with the same biological activity of SEQ ID No.2 polypeptide is a fragment with essentially preserved biological function or activity.

The CCII prepared in this invention is a recombinant protein, polypeptide or fragment, derivative, and analogue. In particular, the polypeptide fragment, variantvariant, or analogue of SEQ ID No.2, might be: (i) a polypeptide with one or several amino acid residues replaced by conservative or non-conservative amino acid residues (a conservative amino acid residue is preferable), where the replaced amino acid residue may or may not be encoded by the genetic code. For example, the mutant or the equivalent of CCII may be obtained by inserting, replacing and/or deleting the amino acid residue. The conservative replacement is based on the similarity in terms of the equivalent charge, solubility, hydrophobicity, and/or amphipathy of the amino acid residues, as long as it maitains the activity of CCII; or (ii) a polypeptide with one or several amino acid residues containing a substituted group; or (iii) a mature polypeptide or other functional compound, such as the polypeptide fused by compounds (for example polyethylene glycol) capable of increasing the half life of the polypeptide; or (iv) a mature polypeptide and polypeptide fusing with

other amino acid sequences which render the amino acid and protein sequence helpful in purifying the mature protein. The structures helpful to purification include NTA mealty affinity chromography, such as the histidine-tryptophan module for purification on the immobilized metal; protein A structure field for purification of the immobilized immunoglobulin, and structure fields for the FLAGS extension/affiliation purification system (IMMUNEX company, Seattle, Wash). The junction sequence specific to XA enterokinase is also helpful in the protein purification (Porath, J. et al. (1992), Prot.Exp.Purif. 3:263-281).

The CCII in the text of the presnet invention comprises the CCII of SEQ ID No.2, i.e., mature polypeptide, the polypeptide having at least 90% similarity to the polypeptide of SEQ ID No.2, 90% identity is preferable), and more preferable, having ate least 95% similarity, and 95% identity is preferable.It also includes part of said polypeptide containing at least 30 amino acids, preferably having more than at least 50 amino acids.

The polypeptide, the conservative variant thereof, and the bioactive fragment and variant thereof, can be prepared by routine synthesis of peptides, such as solid phase synthesis of peptides (Merrifield J. (1963), J. Am. Chem. Soc. 85:2149-2154; Roberge, J. Y. et al. (1995) Science, 269:202-204). The synthesis of the proteins can be performed by hand or by using Applied Biosystems 431A Perkin Elmer. CCII in this invention may also be obtained from the separation and purification of the natural biomaterial, but DNA recomposition using the polynucleotide is preferable.

Based on routine DNA recombination, the polynucleotide sequence of present invention was used to express or prepare the recombinant CCII, thus the present invention relates to the method for producing CCII of the present invention. The method included:

- (1) Transforming the host cell by the polynucleotide (or its variant) encoding CCII, or the recombinant vector containing the polynucleotide.
- (2) Culturing the host cells under proper conditions and in proper culture medium

(3) Separating and purifying the proteins of interest from the culture medium or cells

The present invention also relates to the recombinant vector containing polynucleotide, host cells containing the vector, and the preparation of the polypeptide by recombinant techniques.

The polynucleotide can be used to produce polypeptides using various recombinant techniques. For example, the polynucleotide may exist on any vector selected to express the polypeptide, including chromosomal, non-chromosomal and synthesized DNA sequences, such as the SV40 variant, bacterial plasmid, phage DNA, zymoid plasmid, vectors derived from binding of the plasmid and phage DNA, virus DNA, baculovirus, cowpox virus, adenovirus, fowlpox virus and PRV, and also including, but not restricted to, pQE series (Qiagen), pBS, pD10, phagescript, psiX174, pBluescript SK, pBSKS, pNH series (Stratagene), pTRC99a, pKK223-3, pDR540, pRIT5 (Pharmacia); pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene), pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any plasmid or vector can be used only if they can reproduce and survive in the host.

The present invention also includes the resulting recombinant construct containing the invented polynucleotide, including vectors, such as the above-mentioned plasmid or virus vector, into which the polynucleotide sequence is inserted in the forward or reverse orientation. The construct also contains the regulatory sequence, for example the promoter effectively connected to the polynucleotide sequence (including the constitutive and induced promoter), and mediates the transcription of the down-stream sequence. The promoters include, but are not restricted to: the PL promoter of λ phage, the protein promoter of the baculovirus polyhedron; bacteria promoters, such as LacI, LacZ, T3, T7, gpt, λ PR, PL and trp; eukaryotic promoters, such as CMV early promoter, HSV thymidine kinase promoter, early and late SV40 promoter, retronituse virus LTR and rat metallothioneinI promoter; and also promoters derived from the plant genome such as heat shock protein (HSP) and RUBISCO promoter.

The expression vector contains binding sites for ribosomes and the transcription terminator needed for initiating translation, and also the proper sequence to promote the expression, such as an enhancer, a cis-acting factors for DNA, of about 10-300bp, which stimulates the promoter, for example, an enhancer of 100-270bp after the replication origin in SV40, the enhancer of the early cytomegalovirus promoter, the polyoma enhancer at the posterior side of the replication origin, and the adenovirus enhancer. In addition, the preferable vector also includes one or several selective resistance and/or marker genes that provide phenotypic features to facilitate screening of the transformed host cells; selective genes, such as *trpB* and *hisD* are helpful when using indole or histidinol (Hartman, S. C., and R.C.Mulligan (1988) *Proceedings of the National Academy of Sciences, USA* 85:8047-51), herpes virus TK and APRT for tk- or aprt-cells (Wigler, M. et al. (1977) *Cell* 11:223-32) (Lowy, I. et al. (1980) *Cell* 22:817-23); resistance genes, such as dihydrofolate acid reductase DHFR endow the methotrexate with resistance (Wigler, M. et al(1989), *Proceedings of the National Academy of Sciences, USA*, 77:3567-70) or npt which endows the neomycin and G-418 with resistance (Colbere-Garapin,F. et al. (1981) *J.Mol.Biol.*, 150:1-14), as well as the tetracycline and ampicillin resistance genes. Mammalian expression vectors generally consist of replication origin, promoter, enhancer, and necessary binding sites for ribosomes, polyadenylation sites, the sites for assembling the donor and receptors, transcription termination sequences, and 5' flanked nontranscription sequences. DNA of the SV40-derived assembling sequence and polyadenylation site might provide the necessary nontranscription genetic elements. In addition, vectors containing the polypeptide sequence encoding the polypeptide might also contain specific homologous or heterogenous sequences of the signal peptide, which can aid secretion of the proteins of interest from the procaryotic cells or the eukaryotic cytolemma. Our technicians are aware that the flag sequence and signal peptide in the vectors and construct can be added to the polynucleotide by various recombinant or chemical methods.

A skilled person in the art should be familiar with how to choose a proper vector and

promoter. Vectors for bacteria can be constructed as follows: the DNA sequences encoding the proteins of interest, along with the translation initiating and terminating signals, are inserted into an operable reading frame containing a promoter. It is also well known to a skilled artisan how to construct the nucleotide sequence, as well as the transcription and translation regulatory elements. These methods include *in vitro* DNA recombination, synthesis, and *in vivo* genetic recombination (Sambrook, J. (1989), Molecular cloning: a laboratory manual, Cold Spring Harbor Press; Plainview, N.Y.; Ausubel, F.M. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, N.Y.).

Vectors containing the above-mentioned DNA sequence, promoter, or controlling sequence are suitable to transform the host to express the protein.

It is also within knowledge for a skilled person in the art how to choose a proper host to express the proteins of interest based on the category and the characteristics of the vector. The hosts suitable for expressing polypeptide include, but are not restricted to: a pronucleus host, such as colibacillus, bacillus, streptomyces; a eukaryotic host, such as saccharomyces, aspergillus; insectan cells, such as fruit fly S2 and noctuid Sf9; animal cells such as CHO and COS (Gluzman) (Cell 23:175, 1981) and other cell lines that can express the vectors, such as C127, 3T3, CHO, HeLa, BHK, Bowes melanoma cells; Botanic cells, adenovirus, etc. In addition, various culture systems of mammalian cells are also suitable for expressing recombinant proteins.

The recombinant products were obtained by inserting the vector or the construct containing the above-mentioned nucleotide sequence into the host cells, using the traditional methods, such as calcium chloride mediated transformation, calcium phosphate transfection, DEAE-dextran mediated transfection, electroporation, microinjection, bombard of particles, or using a particle gun, which should be performed by a trained individual. (Sambrook, J.(1989), Molecular cloning: a laboratory manual, Cold Spring Harbor Press; Plainview, N.Y.; Ausubel, F.M. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, N.Y.;

Hobbs, S. et al., McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, N.Y. 191-196; Engelhard, E.K. et al, Proceedings of the National Academy of Sciences, USA, 91:3224-3227; Logan, J. et al, Proceedings of the National Academy of Sciences, USA, 81:3655-3659).

The transformed host strains or cells are cultured to the appropriate cell density, and then the selected promoters are induced using the proper methods, such as altering temperature, and the addition of chemicals. The cells are then further cultured by trained individuals.

Under the control of the proper promoter, the mature proteins can be expressed in mammalian cells, yeast, bacteria, and various other cell types as well. They can also be produced from the DNA construct-derived RNA by using a cell-free translation system. (Sambrook, J. (1989), Molecular cloning: a laboratory manual, the 4th part, 18th chapter, Cold Spring Harbor Press; Plainview, N.Y.)

The cells and the culture solutions are generally harvested by means of centrifugation. When the proteins of interest are contained within the cells, any convenient physical, chemical, or enzymatic methods can be used, including recurrent freezing-melting, ultrasonication, mechanical crush, using cell solvent or specific smudge cells. In this case, the crude extraction is preserved for future purification. When the proteins of interest are secreted out of the cells, routine methods can be used to harvest them directly from the supernatant. The above-mentioned methods should be within the knowledge of a skilled artisan.

The methods to recover and purify the polypeptide from the recombinant cell cultures include ammonium sulfate and ethanol deposition, acid extraction, ion-exchanging chromatography, volume exclusion chromatography, hydrophobic interaction chromatography (HIC), affinity chromatography, hydroxyapatite chromatography, and phytoagglutinin chromatography. Overlap between the proteins is needed in order to allow for them to form their native conformation. Usually, high performance liquid chromatography (HPLC) and

capillary electrophoresis (CE) can be used in the final purification step.

The present invention relates to the pharmaceutical use of the CCII prepared from a polynucleotide sequence in treating and/or preventing RA, including the encoding regions or partial fragment of the nucleic acid molecules, such as full length sequences, partial sequences, and mutant, all of which could effectively express CCII or the functional fragment thereof.

The present invention relates to a pharmaceutical composition for treating and/or preventing osteoarthritis and RA, containing a therapeutically effective amount of CCII prepared according to the methods described in this invention, and optionally, a pharmaceutically accepted vehicle.

The present invention also relates to a food or beverage composition, which is characterized by containing certain amount of CCII prepared according to the method described in this invention. The health food or additive made from CCII will be determined to be beneficial to osteoarthritis and RA.

The present invention relates to a food additive composition containing certain amount of CCII prepared following the methods described in this invention.

The present invention also relates to the usage of the nucleic acid molecules or the fragment thereof in gene therapy.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG 1 illustrates the PCR cloning of the genes encoding CCII in the present invention.

FIG 2 illustrates the structure of the full length CCII in the present invention..

FIG 3 contains PCR results of the 3' untranslated region (UTR) of genes encoding CCII from chicken embryos

FIG 4 illustrates the tissue specificity of the genes encoding CCII.

FIG 5 is a graph showing the standard curve of CCII concentration-absorbancy.

FIG 6 is an analysis of the splicing of the exon of the genes encoding CCII

FIG 7 illustrates the chromosome band of the genes encoding CCII.

FIG 8 is the division phase of the chromosome

FIG 9 shows the results of ISH hybridization.

FIG 10 is a homologous comparison of the evolutionary tree between human being, dog, rat, chicken and speckled dace.

FIG 11 is the western blot analysis of the expression products in the supernatant and cytoplasm induced by pPIC9K/CCOL2A1 (A) and pPICZ α B/CCOL2A1 (B). M is the protein marker, 1-5 is positive band and 6 is blank vector control.

FIG 12 is the WB analysis of the co-expression of pPICZ α B/CCOL2A1/GS115 and pPIC9K/P4H α pPIC9/P4H β . M is the protein marker, 1-5 is positive band and 6 is blank vector control.

The following examples are set forth by way of illustration and not intended to limit the scope of the present invention.

The protocols for the included techniques, including molecular biology, biochemistry and immunology, were all followed from a guide to Molecular Cloning, 2nd edition; digestion conditions for restriction enzymes and buffer were used according to the manufacturers' specifications. The incubation time was 1h at 37°C

unless otherwise stated and in those cases were changed according to manufacturers' instructions.

EXAMPLE 1 — TOTAL RNA EXTRACTION FROM CHICKEN EMBRYO STERNUM

After the eggs had been sterilized, 17-day old SPF chicken embryos (Institute of Animal Sciences, Chinese Academy of Agricultural Sciences) were harvested and washed in cold

saline. The sterum was isolated and grinded in TRIzol reagents on ice, followed by incubation in 1.5 ml eppendorf tube at room temperature for 10min, centrifugation at 1000 x g for 10min at 4°C

. The supernatant was transferred to a DEPC-treated eppendorf tube. To each tube was added 200µl chloroform, followed by extraction mixture, incubation at room temperature for 10min, and centrifugation at 1000 x g for 10min at 4°C

. The aqueous phase was transferred to a new eppendorf tube. To each tube was added the same volume of isopropanol, followed by incubation at room temperature for 10min, centrifugation, washing with 75% ethanol, Rnase-free water dissolving and storage at -80°C for reverse transcription.

cDNA synthesis

cDNA was synthesized from total RNA using a oligodT-3 sites adaptor primer in a 3'RACE kit (TAKARA) as the antisense primier. Following revese transcription, the mixture was incubated at 95°C

for 5 min to inactivate the reverse transcriptase. The cDNA was kept at -20°C for further experiments.

PCR amplification of cDNA encoding CCOL2A1 C-terminal propeptide

According to known CCOL2A1 C-terminal UTR sequences (Sandell LJ, Prentice HL, Kravis D, Upholt WB., J Biol Chem, 1984,259 (12) 7826-7834), sequence encoding of the CCOL2A1 C-terminal propeptide was amplified using col2a1-2F and col2a1-2R primers (Table 1). PCR reactions were carried out in 50 µl final solutions including 1µl (20pmol/µl) of each primer, 10× PCR buffer, 5 µl of 25 mM MgCl₂, 1 µl of 10 mM dNTPs, 2.5 µl of 100% glycerine (final concentration 5%), 2 µl cDNA, and 0.8 µl Taq polymerase. The PCR conditions were 96°C

for 10 min, then 30 cycles of 96°C

for 1 min, 62°C

for 1 min, 72°C

for 3 min and a final extension at 72 °C

for 7 min. The amplified products were separated on a 1% agarose gel, recovered using a gel extraction kit (Invitrogen), cloned into a pGEM-T Easy vector (Promega), transformed into DH5 α and sequenced by the ABI 377 autosequencer.

CCOL2A1 3'UTR cDNA cloning

According to the known CCOL2A1 gene sequence, a CCOL2A1 3'UTR sense primer col2a1-1F (Table 1) was designed and amplified. Using the 3'RACE strategy, the corresponding 3 sites adaptor primer (col2a1-1R) was reverse-transcribed to amplify the 3'UTR. The reaction mixture included the PCR buffer and primer col2a1-1R with 3 site adaptor primers. The PCR conditions were 4 cycles of 96 °C

for 1 min, 62 °C

for 1 min, 72 °C

for 30 s, then 26 cycles of 96 °C

for 30 s, 62 °C

for 30 s, 72 °C

for 20 s and a final extension at 72 °C

for 7 min. Amplified products were recovered and transformed as described above.

C COL2A1 N-propeptide cDNA cloning

The gene encoding the 5'N propeptide was amplified using the sense primer col2a1-5F and antisense primer col2a1-5R. The PCR was performed in PCR buffer, and the annealing and extension steps were conducted at the same temperature. The PCR condition was 30 cycles of 96 °C

for 5 min, 96 °C

for 1 min, 72 °C

for 50 s and a final extension at 72 °C

for 7 min.

Construction and sequencing of recombinant plasmid vector

Plasmid extraction, restriction enzyme digestion, gel extraction, ligation, transformation, construction of recombinant plasmid and identification were performed according to Molecular Cloning, 2nd edition (Sambook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual. 2nd edition Cold Spring Harbor Laboratory Press, 1989). The constructed CCOL2A1 cDNA fragments, SOE ligated CCOL2A1 cDNA fragments and full-length CCOL2A1 were inserted into pGEM-T (Promega). The inserts were identified by restriction enzyme digestion followed by sequencing.

Plasmid construction and preparation

PCR products were recovered using a quick gel extraction kit and were separated by electrophoresis in a 1% agarose gel. Amplified bands were cut out of the gel, and target genes were recovered using a quick gel extraction kit. The process is as follows:

1) The gel containing target gene was put in an eppendorf tube, added 1 ml of L1 solution and incubated in water bath at 50°C

for 15 min, completely dissolving the gel.

2) The dissolved gel was added to a column for centrifugation at 10000 x g for 2 min followed by the addition of 500 µl of L1, incubation at room temperature for 1 min and centrifugation as described above.

3) 700 µl of L2 was added to the mixture, incubated at room temperature for 5 min and centrifuged as described above.

4) The mixture was further centrifuged at 10000 x g for 2 min resulting in volatilizing ethanol.

5) The column was transferred to a new eppendorf tube, followed by the addition of 30 µl of hot (~70°C

) TE buffer, incubation at room temperature for 2 min, and centrifugation at 10000 x g for 2 min. The purified target DNA was passed through into the collection vial. DNA recovered from the gel was separated by electrophoresis on a 1% agarose gel, and the

confirmed concentration was 40 µg/µl.

Ligation of target DNA and pGEMT vector: 10 µl of ligation system containing 5 µl of 2× ligation buffer, 3 µl of target DNA, 1 µl T vector, 1 µl of T4 DNA ligase was mixed and incubated at 4°C

°C overnight.

The competent. E . coli. DH5A was prepared as described by Molecular Cloning, 2nd edition.

The recombinant construct was used to transform DH5α competent cells: 4 µl of the above constructed vectors pGEMT-col2a1-1 ((2, , 3, , 4, , 5 etc.)) were added into 100 µl of freshly prepared competent cells, incubated on ice for 30 min, heat-shocked at 42°C

for 90 s, followed by incubation in ice-water for 2-3 min, addition of 900 µl of LB without ampicillin, shaking at 160 rpm for 45 min at 37°C

, and addition of 4 µl of 1 M IPTG and 16 µl of 50mg/ml X-gal on an LB/Ampicillin agar plate.

The E. coli with that contained the inserts was spread on the plate, incubated at 37°C

for 16 h until a suitable size of colonies appearing.

Screening of positive transformants

White colonies on ampicillin LB plate with IPTG/X-gal were picked up, and inoculated in LB solution with 200 µg/ml ampicillin, incubated with shaking at 37°C (200 rpm) for 12 h, followed by plasmid extraction.

STEP 6: Screen a selected few white colonies.

Using above described method, the present invention produced pGEMT-col2a1-2, pGEMT- col2a1-1, pGEMT- col2a1-3, pGEMT- col2a1-(1+2), pGEMT- col2a1-5 and pGEMT- col2a1-4.

The positive colonies were identified by cutting the plasmid in 20µl of digestion system containing 1 µl of 10× Not I and Nco I restriction enzymes, 1 µl of 0.1% BSA, 8µl of

plasmid and 8 μ l of water using a Promega A7500 kit. The mixture was incubated at 37°C for 2 h, followed by electrophoresis on a 1% agarose gel and sequencing by the ABI 377 autosequencer.

Splice overlap extension-polymerase chain reaction (SOE-PCR) of full-length chicken CCOL2A1 cDNA

In order to obtain full-length CCOL2A1 cDNA, this research the described research used a SOE-PCR strategy for the construction (Horton RM, Hunt HD, Ho SN, et al. , , Gene, 1989, 77:61-68). This procedure needed four-times required four phases of PCR.

The first PCR reaction used pGEMT-col2a1-2, pGEMT- col2a1-1 PCR products as templates, col2a-2F and col2a1-1R as sense and antisense primers and Pfu DNA polymerase to link products together resulting in products col2a1- (2+1). The PCR conditions were 4 cycles of 96°C

for 5 min, 96°C

for 1 min, 60°C

for 1 min and 72°C

for 150s, then 26 cycles of 96°C

for 30 s, 58°C

for 30 s, 72°C

for 130 s and a final extension at 72°C

for 7 min.

The second PCR used plasmids pGEMT- col2a1-3, pGEMT- col2a1-(1+2) PCR as templates, col2a-3F and col2a1-1R as sense and antisense primers. The PCR conditions were 4 cycles of 96°C

for 5 min, 96°C

for 1 min, 64°C

for 1 min, 72°C

for 120 s then 26 cycles of 96°C

for 30 s, 62°C

for 1 min, 72°C

for 90 s and a final extension at 72°C

for 7min.

The third PCR used col2a-5F and col2a1-4R as sense and antisense primers, col2a1-5 and col2a1-4 plasmid PCR products as templates. The PCR conditions were 4 cycles of 96°C

for 5 min, 96°C

for 1 min, 62°C

for 2 min, 72°C

for 120s, then 26 cycles at 96°C

for 1 min, 60°C

for 1 min, 72°C

for 90 s and a final extension at 72°C

for 7 min.

The 4th PCR used col2a-5F and col2a1-1R as sense and antisense primers, former 3-time PCR-linked PCR products (col2a1-5+4, col2a1 1+2+3) as templates, resulting in full-length col2a1 cDNA. PCR conditions were 4 cycles of 96°C

for 5 min, 96°C

for 1 min, 62°C

for 2 min, 72°C

for 210 s, then 26 cycles at 96°C

for 1 min, 62°C

for 1 min, 72°C

for 280 s and a final extension at 72°C

for 7 min

The primers used for cloning full-length chicken CCOL2A1 cDNA are shown in Table 1

Table1. Primer sequences for amplifying full-length chicken CCOL2A1 cDNA

col2a1-1F	5'- TCT ATC GCG CAC CCG TTG TGC -3'
col2a1-1R	5'- GTC TTG TAG TGC TAC GGC TTG C -3'
col2a-12F	5'- TTG CAG ATG TCT CCA ATA CCA G -3'
col2a1-2R	5'- GCA CAA CGG CTC GGG CAA TGT GCT AAC G -3'
col2a1-3F	5'- GCT CGG AAG CAA CGG CCT CG -3'
col2a1-3R	5'- CTC GTC CCG GAC GCG ACG G -3'
col2a1-4F	5'- CGC TGC GAT CGT CAT GCG G -3'
col2a1-4R	5'- GTA GTG ACC CTA CGC CCG AG -3'
col2a1-5F	5'- ACG CCG GCT CTC GTG CTC CTC GTG GTG C -3'
col2a1-5R	5'- CCG CCC GGG TCC GAA TGC CCG CAT -3'

The present invention. utilized a SOE-PCR strategy to obtain full-length chicken CCOL2A1 cDNA. Due to the high GC content in the chicken CCOL2A1 gene, the buffer suitable for amplifying high GC content was used in the reaction system during the amplification procedure. The amplified products possessed high specificity and were inserted into pGEM-T vectors. The vectors were replicated in *E. coli* DH5, digested by the restriction enzymes provided by Promega, transformed and sequenced. The primers described above were synthesized by Shanghai Bioasia Biotech Co, and sequencing was performed by Takara Co. (Dalian, China). The results are shown in the sequencing list in SEQ ID NO:1.

EXAMPLE2 - PREPARATION CHICKEN GENOMIC DNA

Chicken genomic DNA was extracted using a Wizard genomic DNA purification kit (Promega) according to the manufacturer's instructions. The purity of the extracted DNA was 1.6-1.8 of 260/280 OD value using a spectrophotometer (BECKMAN, DU®640).

PCR-based cloning and library screening of chicken CCOL2A1 genomic DNA

PgF and PgR were used as sense and antisense primers for CCOL2A1 C-terminal cloning. The sequences of primers were as follows: PgF 5' CCA GGC AAG GAT GGC GCA CG 3'; PgR 5'CCT GAT CGG CTC CGC CAA TGT CCA TAG G 3'. PCR of CCOL2A1 genomic DNA was performed using LA Taq polymerase in GC buffer. PCR products were examined by 0.8% agarose gel electrophoresis, target gene fragments were recovered from gel slices using a gel extraction kit, and inserted in pGEM-T vectors. The inserts were identified by restriction enzyme digestion and sequencing. Due to the high GC content in the CCOL2A1 genomic DNA and polyT structure, fragment cloning of N-terminal of CCOL2A1 genomic sequences were performed by the SOE-PCR method using the total DNA extracted from whole blood and spliced to generate a nearly full-length CCOL2A1 genomic DNA sequence.

Initially, we amplified the DNA extracted from peripheral chicken blood via PCR to obtain a 5494bp fragment at the CCOL2A1 3' end, and the specificity of the amplified products was very high (Figure 1-9). The fragment was inserted into a pGEM-T vector to get pGEM-T/CCOL2A1 and confirmed by sequencing.

The results of sequencing revealed that the 3' ends of the chicken CCOL2A1 genomic DNA contained a partial triple helix-forming region, 3'-terminal peptide and propeptide. The cloned genomic regions contained 19 introns and 20 exons.

However, of the genomic sequences beyond the 3' end, regardless of templates, primers and PCR conditions used, the target gene sequences could not be obtained. Afterwards, the positive colonies screened from the genomic chicken DNA library were sequenced, but high GC content and PolyT structure restricted the process of sequencing, resulting in a sequencing signal interrupt.

Therefore, we utilized known exon sequences to design numerous primers to complete the CCOL2A1 genomic DNA sequencing. Finally, we obtained a 12523bp length CCOL2A1 genomic DNA sequence containing 54 exons and 53 introns. The exons and introns are shown

in Table 2. The structure of the obtained chicken COL2A1 3' end gene is shown in Figure 2. After accessing the database, it was clear that the genomic DNA sequences specifically belonged to the COL2A1 genome, and had not yet been reported. The relevant gene sequences have been submitted to GenBank (Accession No. AF52711).

Table 2. The structure of exons and introns of the 3' genomic chicken COL2A1 DNA

Exon	Amino acid coding region	Exon sizes (bp)	Intron sizes (bp)
1	1-30	>135	ND
2	31-100	210	ND
3	101-117	50	ND
4	118-146	87	ND
5	147-180	102	ND
6	181-206	78	ND
7	207-221	45	ND
8	222-239	54	ND
9	240-258	54	124
10	259-274	54	107
11	275-292	54	100
12	293-310	54	98
13	311-325	45	86
14	326-343	54	110
15	344-358	45	86
16	359-376	54	113
17	377-409	99	89
18	410-424	45	198
19	425-457	99	367
20	458-475	54	885
21	476-511	108	408
22	512-529	54	88
23	530-562	99	91
24	563-580	54	682
25	581-613	99	82
26	614-631	54	82
27	632-649	54	124
28	650-667	54	277
29	668-685	54	79
30	686-700	45	114
31	701-733	99	262

32	734-769	108	92
33	770-787	54	80
34	789-805	54	81
35	806-823	54	82
36	824-841	54	94
37	842-877	108	89
38	878-895	54	108
39	896-913	54	113
40	914-967	162	268
41	968-1003	108	92
42	1004-1039	108	309
43	1040-1057	54	160
44	1058-1093	108	97
45	1094-1111	54	268
46	1112-1147	108	109
47	1148-1165	54	134
48	1156-1201	108	79
49	1202-1297	289	78
50	1298-1360	188	407
51	1361-1441	243	594
52	1442-1489+	667	112
520bp 3'UTR			

Notes: Exon and introns were named according to the method described by Upholt et al. (J Biol Chem, 1984, 259(12); 826-7834).

EXAMPLE 3 CHICKEN CCOL2A1 α_1 (II) SPECIALLY EXPRESSION IN TISSUES

CCOL2A1 expression spectrum in the chicken embryo

CCOL2A1 3' UTR was confirmed to specially belong to CCOL2A1 via Blast database searching. Therefore, the CCOL2A1 cDNA 3'UTR was used as an amplifying subject to investigate α_1 (II) expression in the developing chicken embryo. Heart, liver, vitreum, cornea, skin, pectoralis, sternum and etc. in 17-day old chicken embryo were analyzed using RT-PCR. The primers were col2a1-1F and col2a1-1R, listed in Table 1. The PCR products were inserted into a pGEMT-easy vector and sequenced after restriction enzyme digestion. Simultaneously, GAPDH was amplified as an internal control using the

primers PF_{GAPDH} 5' GC AGA GGT GCT GCC CAG AAC 3'; PR_{GAPDH} 5' TCA CTC CTT GGA TGC CAT GTG 3' to generate a 412bp GAPDH fragment.

Using the results obtained from the RT-PCR of 17-day-old chicken embryo CCOL2A1 3' UTR, chicken embryo CCOL2A1 mRNA was expressed in heart, liver, vitreum, cornea, skin, pectoralis, sternum, small intestine, arthrodial cartilage, meniscus and cranial bone, but was not detected in spleen, thymus and testis (Figure 3).

Expression spectrum of Adult chicken type II collagen

Type II collagen in heart, liver, spleen, kidney, vitreum, cornea, skin, pectoralis, meniscus, pancreas, thymus, small intestine, stomach, testis, skeletal muscle, cerebrum, cerebellum, arthrodial cartilage, sternum and lung of 4-week-old chicken was analyzed by ELISA using a natural type II collagen test kit (Chondrex). Five mg of each frozen sample was homogenized and dissolved in 0.8 ml of 50 mM acetic acid-0.2 M NaCl (pH 2.9-3.0), digested in 100 µl of 20 mg/ml pepsase (Sigma) at 4°C

for 48h followed by addition of 250 µl of 10×TSB. The pH was adjusted to 8.0 using M NaOH, samples were incubated in 100 (1 of 2 mg/ml elastase (Sigma) at 4°C °C overnight, centrifugation and centrifuged at 10000 x g for 5 min at 4°C °C. The precipitate was discarded, the volume of the supernatant was adjusted to 1ml and, discharging precipitation, adjusting the supernatant to 1 ml and stocked at 4°C

for future experiments.

The extracted chicken type II collagen was analyzed by sandwich ELISA. Two monoclonal antibodies against different epitopes of the chicken type II collagen molecule were used. One was used as a coating antibody, and the other was used as the enzyme-labeled detecting antibody. First, an ELISA plate was coated with the capture antibody at 4°C

overnight, followed by 3 washes, 1:200 dilutions of the samples, loading samples and

standards onto the coated ELISA plate, incubation at 37°C

for 2 h, washing as described above, adding biotin-labeled anti-chicken type II collagen monoclonal antibody to each well, incubation at 37°C

for 2 h, washing as described above, adding enzyme-conjugated streptavidin, incubation at 37°C

for 1 h, washing vigorously, adding OPD solution at 37°C

for color development, stopping the enzyme reaction by 1.25 M sulfuric acid and analyzing the OD490 value. The detailed protocol was described in the kit's instructions (The Arthrogen-CIA® native type II collagen detection kit (Chondrex)).

The results of the natural type II collagen analysis in 20 organs of adult chickens showed that chicken type II collagen is expressed in sternum, arthroal cartilage, pancreas and small intestine, but not in other organs (Figure 4).

Using chicken natural type II collagen as a standard, a concentration vs absorbance standard curve was made to calculate the expression level of chicken type II collagen in relevant positive tissues (Figure 5).

Based on the standard curve, the expression level of chicken type II collagen was 0.4% in sternum, 1.5% (the highest level in arthrodial cartilage of 4-week chicken), 0.84% in pancreas and 0.2% in small intestine.

EXAMPLE 4 - SPLICING ANALYSIS OF CHICKEN CCOL2A1 5' PROPEPTIDE EXON

2

Splicing analysis of the CCOL2A1 exon 2 in the tissue from chicken embryos was performed using RT-PCR. mRNA extracted from skin, liver, vitreum, cornea, pectoralis, small intestine, sternum, arthroal cartilage and etc., and col2a1-5 and col2a1-5R sense and antisense primers listed in Table 1 were used for RT-PCR analysis. The specificity of the

PCR products was determined by 1% agarose gel electrophoresis. The results of RT-PCR showed that exon 2 of chicken CCOL2A1 5' N propeptide was spliced and transcribed in heart, liver vitreum, cornea, pectoralis and small intestine, but not in cartilaginous tissue such as sternum and arthroal cartilage (Figure 6).

EXAMPLE 5-MAPPING CHICKEN CCOL2A1 GENE CHROMOSOMES

In order to precisely map map CCOL2A1 to the chromosome, using *in situ* hybridization (ISH) and DIG-High Prime DNA Labeling and Detection Starter Kit II (Roche), the gene fragments of the 3'-UTR chicken CCOL2A1 gene were labeled by digoxin as probes, and hybridized with the metaphase chromosome to map the chromosomes. After statistical analysis, the chicken CCOL2A1 gene was located to chromosome 4p2 (Figures 7, 8 and 9).

EXAMPLE 6- CHICKEN CCOL2A1 HOMOLOGICAL COMPARISON

The homology of the cDNA and amino acid sequence of the type II collagen triple helix-forming region of the chicken CCOL2A1 was compared with that of various species such as dog (AF023169, AF242201), human (MM001844), zebrafish (U23822), and mouse (M65161) searched from GenBank using DNASTAR Megalign software (DNASTAR, Inc.). A phylogenetic tree was drawn using the Genedoc program (Figure 10).

The results indicate that the amino acid and cDNA sequence of chicken CCOL2A1 had 79.03% and 94.77% homology respectively with dog CII suggesting the highest homology in phylogeny; the next was between chicken and human, showing 78.96% and 93.89% identity. The homology between human and dog showed the highest identity among these five species, showing 91.89% identity of amino acid and 98.52% identity of cDNA, the next was between human and mouse, showing 88.63% and 96.15% identity respectively. A phylogenetic tree was constructed to represent the comparison results and analyze the

homology.

EXAMPLE 7 - EXPRESSION OF RECOMBINANT CCOL2A1 cDNA

The materials, reagents and instruments used in this experiment are described below:

The expression vectors pPIC9K, pPIC9, pPICZ α A, B, C and yeast strains *Pichia pastoris* GS115 and X33, Spheroplast kit for yeast and *Pichia* EasyCompTM were purchased from Invitrogen; Mung bean nuclease was from NEB; Mini protein 2 electrophoresis system and semi-dry electrophoretic transfer were from Bio-Rad; Gel-Pro image analyser system V3.1 was from Media Cybernetics; DU640 nucleic acid and protein analyzer was from Beckman.

1) The construction of the eukaryotic expression vector pPICZ α B/ CCOL2A1

The eukaryotic expression vector was constructed as follows: PCR was performed using the primers (sequence 25 5'GGT ACC TTG GTG GAA ACT TTG CGG 3') and (sequence 26 5'GGT ACC GTT ACA AGA AGC AGA CTG 3'), and pGEM-T/CCOL2A1 as template. The PCR mixture contained 25 μ l of 2 \times GC Buffer I, 4 μ l of dNTP (2.5 mM), 0.5 μ l of 20 μ M primers. The PCR condition was after 96°C

at 5 min, then 4 cycles at 96°C

for 1 min, 6°C

for 1 min, 72°C

for 3 min and 26 cycles at 96°C

for 1 min, 65°C

for 30 s, 72°C

for 2.5 min and a final extension 72°C

for 5 min. The recovered PCR products were inserted into a pPICZ α B expression vector, followed by a determination of the correct direction of inserts in expression vector pPICZ α B/ CCOL2A1 by *Bam*HI digestion.

2) Construction of yeast expression vector pPIC9K/CCOL2A1

In order to improve the stability and quantity of expression of the pPICZ α B/CCOL2A1

vector in yeast *P. pastoris*, the CCOL2A1 fragment was inserted into the high copy vector pPIC9K, resulting in pPIC9K/CCOL2A1. The construction procedure is shown in Figure below. Briefly, pPICZαB/CCOL2A1 was linearized by digestion with *Kpn* I then digested by mung bean nuclease to generate blunt-ends. Simultaneously, pPIC9K was digested by the blunt-end enzyme *Sna*BI. The two fragments were ligated by T4 DNA ligase, resulting in an expression vector pPIC9K /CCOL2A1, and open reading frames were confirmed by sequencing

3) *P.pastoris* GS115 preparation protoplast transformation method

The yeasts were transformed using theprotoplast method:

a) *P. pastoris* GS115 was spread on YPD plates by streak inoculation, followed by incubation at 30°C

for 2 d, inoculation of selected colonies in 5 ml of YPD medium, and incubation at 30°C with agitation at 300rpm overnight

b) The cell suspensions with 1:1000 dilution was inoculated in 100ml of YPD, incubated at 30°C

and agitated at 300 rpm until OD600 reached 0.2-0.3, followed by centrifuging at 1,500 x g for 5 min and discharging the supernatant.

c) The cells were washed once with 20 ml sterilized water, 20 ml SED and 20 ml 1M sorbitol, followed by centrifugation at 1500 x g for 5 min. The supernatant was discarded after each wash. The cells were resuspended in 20 ml of SCE, and divided 10 ml each into A and B tubes.

d) 800 µl of 15%SDS was added to each 1.5 ml eppendorf tube labeled 0, 2, 4, 5, 6, 8, 9, 10, 15, 18, 20, 25, 30, 35, 40, representing the time in the Zymolyase digesting cell well (min). 200µl cells suspensions were taken from the A tube and frozen for the control of Zymolyase digesting time. 7.5 µl of Zymolyase (2.25 units) was added to the A tube, mixed slightly and maintained at 30°C

. 200 µl of digested cells were taken from the A tube at each time according to the label and mixed with 800 µl of 5%SDS in eppendorf tubes, and placed on ice to stop the enzyme

reaction.

e) The percentage of protoplast forming at each time point was calculated using the following equation: $\text{protoplast \%} = 100 - [(\text{OD}_{800}(\text{min})/\text{OD}_{800}(0\text{min}) \times 100)$. This equation was used to calculate the time (t, min) at which 70% of the protoplast was formed. 7.5 µl of Zymolyase was added to the B tube, and maintained at 30°C

for t time.

f) The protoplasts were centrifuged at 750 x g for 10 min. Supernatants were discarded, and protoplasts were washed with 10ml of 1 M of sorbitol and 10ml of CaS respectively, supernatant was discarded, and the protoplasts were re-suspended with 0.6 ml of CaS for transformation.

4) pPIC9K/CCOL2A1 and pPICZαB/CCOL2A1 transformed GS115 protoplast

a) Approximately 10 µg of linearized pPICZαB/CCOL2A1 and pPIC9K/CCOL2A1 were mixed gently with 100 µl of freshly prepared protoplast followed by incubation at room temperature for 10 min. pPIC9K/GS115 and pPICZαB/GS11 blank vectors were used as negative controls.

b) One ml of freshly made PEG/CaT (1:1) solution was added and mixed gently followed by incubation at room temperature for 10 min.

c) Samples were centrifuged at 750 x g for 10 min and as much of the supernatant as possible was removed. Cells were resuspended in 150 µl of SOS and incubated for 20 min, after which 850 µl of 1 M sorbitol was added.

d) Each 150 µl of transformed protoplasts were mixed with 10 ml of RD agar medium, spread on upper RDB plate, incubated at 28-30°C

for 4-6 d, after which the results were obtained.

5) pPICZαB/CCOL2A1/GS115 and pPIC9K/CCOL2A1/GS115 positive transformant screening

Cultured pPIC9K/CCOL2A1/GS115 and pPICZαB/CCOL2A1/GS115 grew approximately 100 transformants on RDB plates. In order to obtain a high copy number, each transformant was transferred to a plate containing various concentration of G418 or

Zeocin (G418 concentration: 0.25 mg/ml, 0.5 mg/ml, 1 mg/ml, 2.0 mg/ml, 3.0 mg/ml, 4.0 mg/ml; Zeocin concentration: 100 µg/ml, 200 µg/ml) to screen for high copy number.

6) Extraction of the yeast genomic DNA, and identification of the integrated target gene in the yeast genome by PCR

Resistant transformants and blank vectors were inoculated into 2 ml of YPD and cultured at 30°C

and shaken at 300rpm for 2d. The genomic DNA was extracted and inserts were detected by PCR using CCOL2A1 cDNA as specific primers: F:5'GGTACC TTG GTG GAA ACT TTG CGG 3', R: 5'GGTACC GTT ACA AGA AGC AGA CTG 3'. PCR conditions were 96°C

for 5min, then 5 cycles at 96°C

for 1 min, 60°C

for 2 min, 72°C

for 3 min, and 25 cycles of 96°C

for 1 min, 58°C

for 1 min, 72°C

for 2 min, and a final extension at 72°C

for 5 min.

7) pPICZαB/CCOL2A1/GS115 and pPIC9K/CCOL2A1/GS115 methanol-induced expression and Western blot analysis

Positively-identified colonies were inoculated in 30 ml of BMGY medium and incubated at 28°C

and shaken at 300 rpm for 2 d. The cells were collected by centrifugation and resuspended with 1/10 volume of BMMY followed by further incubation at 28°C

and shaken at 300 rpm. Methanol was added to 1% of volume at 24-h intervals. The supernatant and yeast were harvested from the medium induced for 4 d. The supernatant was dialyzed, desalinated and lyophilized. The cell lysates from yeast were prepared by

vortexing with 1/4 volume of glass beads followed by collection via centrifugation. The induced cell lysates and supernatants were separated by 10% SDS-PAGE electrophoresis and transferred to nitrocellulose filters. The filters were blocked with 20% FBS at 4°C

overnight, incubated with a monoclonal antibody to 95D1A (1:1000) at room temperature for 2h, and washed 3 times with TTBS, each 5 min. Filters were incubated with goat anti-mouse AP conjugated 2°C

antibody at room temperature for 2 h, and were washed again as described above. The signal was detected using a NBT/BCIP Reagent Kit

Results:

The supernatant from each induced expression vector described above was desalinated, concentrated and identified by 10% SDS-PAGE electrophoresis and Western blot analysis. The results reveal that an 80KD band representing the CCOL2A1 α chain was only in the cytoplasm of the pPICZ α B/CCOL2A1 transformant (Figure 11A); however complete expression of the CCOL2A1 α chain was detected in the cytoplasm of pPIC9K/CCOL2A (Figure 11B), but not in the supernatant by western blot analysis.

EXAMPLE 8 - CONSTRUCTION OF PROLYL 4-HYDROXYLASE α AND/OR β SUBUNIT

1) prolyl 4-hydroxylase α subunit (P4H α) cDNA cloning

According to the chicken P4H α gene sequence, primer a: 5'AGA TAC TGC TAC GAA AGA CCC CGA G 3' and b: 5' CTC TCT TGG TTG TAG CCC TCA TCT G 3' were designed. The PCR system contained 5 μ l of 10 \times PCR Buffer, 5 μ l of 25 mM MgCl₂, 4 μ l of 2.5 mM dNTP Mix, 0.5 μ l of 20 μ M primers, 1 μ l of cDNA ,0.5 μ l Taq, and water to a final volume of 50 μ l. PCR products were examined by 1% agarose electrophoresis. PCR conditions were 96°C

for 5 min, then 4 cycles of 96°C

for 1 min, 72°C

for 3 min and 26 cycles at 96°C

for 1 min, 69°C

for 1 min, 72°C

for 2 min and a final extension at 72°C

for 3 min.

2) Construction of the pPIC9K/P4H α eukaryotic expression vector

The P4H α PCR product described above was ligated with pGEM-T, resulting in a pGEM-T/P4H α plasmid. A *Not* I restriction site was introduced by PCR. The primers a: 5'GCG GCC GCA GAT ACT GCT ACG AAA G 3' and b: 5'GCG GCC GCC TCT CTT GGT TGT AGG 3', and Pfu DNA polymerase were used for PCR. PCR products were given polyA tails then ligated with a pGEM-T vector and sequenced. The target fragments were recovered after the pGEM-T/P4H α was digested by *Not* I. The P4H α gene fragment was ligated with the pPIC9K vector linearized by *Not* I using T4 DNA ligase resulting in a pPIC9K/P4H α expression vector. The direction of insert was identified.

3) P4H β cDNA cloning and recombinant expression vector construction

According to the known chicken P4H β subunit gene sequence, the primers F:5'GCG GCC GCA CAG CCC CTG GAG GAG 3' and R: 5'GCG GCC GCG GTG ATG TAG ATC AGT C 3' were designed and *Not* I restriction sites were introduced. RNA was extracted from sternum of 17-day-old chicken embryos followed by RT-PCR. 1.5 mM [Mg²⁺] was used in the PCR system. PCR conditions were 96°C

for 5 min, then 4 cycles of 96°C

for 1 min, 53°C

for 1 min, 72°C

for 2 min and 26 cycles at 96°C

for 30 s, 51°C

for 30 s, 72°C

for 2 min and a final extension at 72°C

for 5 min. The PCR products were inserted into a pGEM-T for sequencing. The

colonies with the correct sequence were inserted into a pPIC9 expression vector, resulting in the pPIC9/P4H β expression vector.

EXAMPLE 9 - COEXPRESSION OF pPIC9K/P4H α , pPIC9/P4H β AND pPICZ α B/CCOL2A1 IN YEAST GS115

The expression vectors pPIC9K/P4H α , pPIC9/P4H β and pPICZ α B/CCOL2A1 described in Examples 7 and 8 were linearized by *Bgl* II, *Sal* I and *Pme* I. 10 μ g of linearized vector was transformed into GS115. The transformants were cultured on MM plates containing G418 and Zeocin at 30°C

for 4-6d.

EXAMPLE 10 - CONSTRUCTION OF DOUBLE EXPRESSION VECTOR pPICZ α A/P4H α - β

During co-expression of pPIC9K/P4H α , pPIC9/P4H β and pPICZ α B/CCOL2A1, the positive clones were unstable. In order to overcome the deficiency of the construction described above, P4H α and P4H β were constructed into the same expression vector pPICZ α A, resulting in a pPICZ α A/P4H α -P4H β double expression vector.

PCR was performed using the plasmids pGEM-T/P4H α and pGEM-T/P4H β in Example 8 and 9, as templates. The primers' sequences were P4H α : F: 5'GCGGCCGC GAT ACT GCT ACG AAA G3'; R: 5'GCGGCCGC CTC CAA CTC TGA TAA C 3'; P4H β : F: 5'GCGGCCGC CAG CCC CTG GAG GAG -3'; and R: 5' GCGGCCGC TTA ATC ATC ATC AGC 3'. PCR conditions were 96°C

for 5 min, then 4 cycles at 96°C

for 1 min, 66°C

for 80 s, 72°C

for 90 s and 26 cycles at 96°C

for 1 min, 64°C

for 40 s, 72°C

for 50 s and a final extension at 72°C

for 5 min. PCR products were examined by 1% agarose gel electrophoresis. The specific target fragments were recovered and inserted into pGEM-Teasy followed by sequencing.

pGEM-T/P4H α and pGEM-T/P4H β confirmed by sequencing were digested by *Not* I, ligated with pPICZ α A, resulting in recombinant expression vectors pPICZ α A/P4H α and pPICZ α A/P4H β .

A P4H β full-expression unit was cut from pPICZ α A/P4H β , and inserted into pPICZ α A/P4H α . Since *Bgl* II and *Bam* HI are isoschizomers, these sites can link with linearized pPICZ α A/P4H α by *Bgl* II resulting in the double expression vector pPICZ α A/P4H α - β .

Since the P4H β expression unit was located upstream, PCR was used to examine whether the direction of P4H α and β was coincident. The upstream primer (5' GCGGCCGC CAG CCC CTG GAG GAG 3') and downstream primer (5' GCGGCCGC CTC CAA CTC TGA TAA C 3') for amplifying P4H α was used to PCR. If the directions of inserts were correct, 4644bp PCR products would be apparent, otherwise, there would be no expression band.

EXAMPLE 11 - CO-EXPRESSION OF pPIC9K/CCOL2A1 AND pPICZ α A/P4H α - β

pPIC9K/CCOL2A1 and pPICZ α A/P4H α - β were co-transformed into GS115 to study expression of the three genes. Protoplast preparation, positive transformant screening, PCR identifying and methanol induction were performed as described above.

The induced pPICZ(B/CCOL2A1/GS115 and pPIC9K/P4H β) co-expression products in supernatants and cytolysates were identified by SDS-PAGE and Western blot analysis. Figure 12 shows that the full-length of CCOL2A1 (chain was only expressed in the cytolysate, but not in the supernatants).